



*Facultad
de
Ciencias*

**LEYES DE POTENCIAS EN LA ECONOMÍA Y EN LA
EMPRESA**
(POWER LAWS IN ECONOMICS AND BUSINESS)

Trabajo de Fin de Grado
para acceder al

GRADO EN MATEMÁTICAS

Autor: Diego Bergareche Solana
Director: Mario Alfredo Fioravanti Villanueva
Co-Director: Faustino Prieto Mendoza
Junio - 2021

RESUMEN

Dentro del ámbito de las leyes de potencias (distribución clásica de Pareto) se han desarrollado tanto el análisis de las técnicas más adecuadas para su estimación y contraste como su utilización en la modelización matemática de fenómenos no lineales del mundo real en la cola alta de la distribución. Este trabajo consta de dos partes bien diferenciadas. En la primera, se ha implementado un programa en RStudio que permite el análisis de datos empíricos siguiendo la teoría expuesta en Clauset et al. (2009). En la segunda, y junto con el software creado, se han estudiado variables económicas atendiendo a dos propiedades, el comportamiento autosimilar y la economía de superestrellas.

Palabras clave: Leyes de potencias, fractales, economía de superestrellas, distribuciones con colas pesadas.

ABSTRACT

Within the study of Power Laws (classic Pareto distributions) two topics have been of continuous review, the methods used to estimate its parameters for given empirical data and the study of new variables whose tail distributions do follow Power Laws. This work has two main parts. In the first one, an RStudio program was developed to analyze empirical data following Clauset et al. (2009). In the second one, and using the software created, we studied economic indicators under two categories: self-similar ones and those related to the economics of superstars.

Key words: Power Laws, fractals, economics of superstars, heavy tail distributions.

AGRADECIMIENTOS

Para llegar hasta la finalización de este trabajo muchas personas han tenido un impacto indispensable. Habría que empezar por todos aquellos profesores que me inculcaron la semilla de la curiosidad matemática, de analizar, entender y dar valor a los problemas. Son ellos los que me impulsaron a adentrarme en el mundo de la ciencia. Después, agradecer a mi familia, apoyo incalculable durante todo este tiempo, conocedores del esfuerzo que ha conllevado entrar y posteriormente completar ambos grados. Por último, agradecer tanto a Mario como a Faustino su implicación en el proyecto. Al primero por permitirme salirme de la norma y escoger un tema ajeno a la facultad. Al segundo por no haber dudado ni un segundo en trabajar conmigo desde el momento en que mostré interés, *going the extra mile* cuando era necesario.

Índice de contenidos

1.	INTRODUCCIÓN	9
1.1.	LEYES DE POTENCIAS	9
1.2.	MÉTODO DE ANÁLISIS DE LEYES DE POTENCIAS.....	11
	Procedimiento.....	11
1.3.	LEYES DE POTENCIAS EN LA ECONOMÍA	16
	Fractales.....	16
	La economía de las superestrellas	17
2.	METODOLOGÍA DE ESTUDIO	19
2.1.	OBTENCIÓN DE LOS DATOS.....	19
2.2.	ANÁLISIS DE LOS DATOS.....	20
2.3.	LISTADO DE VARIABLES	21
	Variables dentro de la teoría fractal	21
	Variables dentro de la economía de superestrellas.....	22
3.	RESULTADOS	25
3.1.	TABLAS	25
	Comprobación del algoritmo	25
	Variables dentro de la teoría fractal	26
	Variables en la economía de superestrellas	27
3.2.	GRÁFICOS	28
	Comprobación del algoritmo	28
	Variables dentro de la teoría fractal	29
	Variables en la economía de superestrellas	30
4.	ANÁLISIS.....	31
4.1.	ANÁLISIS DEL PROGRAMA	31
4.2.	ANÁLISIS DE LAS VARIABLES ALEATORIAS	31
5.	CONCLUSIONES	33
5.1.	SUGERENCIAS Y MEJORAS.....	34
6.	REFERENCIAS.....	35
7.	APÉNDICE.....	39

1. INTRODUCCIÓN

Cuando uno se dispone a realizar el estudio de la altura de una población adulta, digamos la española, aprecia que los ciudadanos miden de media 1.7 m. Además, no se encontrará con ciudadanos adultos que midan tres metros o diez centímetros, siendo igual de improbable tanto lo uno como lo otro.

Hay otros sucesos que no se comportan de la misma manera. Si miramos a variables tales como la distribución de las poblaciones en ciudades, la fuerza de los terremotos y las variaciones diarias de la bolsa vemos que los eventos más grandes lo son en varios órdenes de magnitud en comparación con aquellos más comunes. Se dice que este tipo de variables siguen leyes de potencias.

Las leyes de potencias han recibido gran atención en las últimas décadas ya que permiten cuantificar fenómenos complejos mediante una elegante y simple fórmula. Además, numerosos modelos han sido creado en disciplinas muy diversas, desde la física a la economía pasando por la biología. Por lo tanto, son una herramienta muy poderosa para la descripción de fenómenos hasta hace poco no comprendidos.

Los artículos publicados hasta la fecha en lo concerniente a este tema se centran principalmente en dos aspectos: o bien encontrar y modelar variables que sigan leyes de potencias, o bien definir un método riguroso para decidir si una variable es una ley de potencias, y en caso afirmativo, qué parámetros son los que la ajustan.

En este trabajo pretendemos aunar ambas facetas. En primer lugar, estudiaremos e implementaremos el método mayoritariamente aceptado para analizar leyes de potencias, expuesto de manera rigurosa en Clauset et al. (2009). Revisaremos los conceptos fundamentales y nos adentraremos en los pasos a seguir. Con ello, se creará un software propio en R que será el que se empleará posteriormente para el análisis de datos empíricos. En segundo lugar, nos adentraremos en el concepto de fractal, y cómo ciertos comportamientos propios de los mismos pueden estar acompañados por una descripción mediante leyes de potencias. Esto nos servirá para estudiar variables económicas mayoritariamente aceptadas como leyes de potencias, analizándolas a diferentes niveles, y viendo si muestran características fractales a través del concepto de autosimilitud. Para terminar, veremos la economía de superestrellas dentro del mundo deportivo. Estudiaremos las variables salario y premios acumulados en torneos para diferentes disciplinas con la intención de ver si se comportan como leyes de potencias. En el caso de los premios acumulados, haremos un desglose entre hombres y mujeres.

1.1. LEYES DE POTENCIAS

En el libro *Course d'Économie Politique*, Pareto (1896), destacó que aproximadamente el ochenta por ciento de la tierra en Italia era propiedad de un veinte por ciento de la población. Este tipo de comportamientos, donde la mayoría de los acontecimientos son resultado de una minoría de causas, pueden ser descritos por leyes de potencias. Genéricamente son expresadas por una función de densidad de la forma

$$f(x) \propto x^{-\alpha}. \quad (1)$$

Al parámetro α se lo conoce habitualmente como parámetro de forma. De aquí en adelante nos referiremos a las leyes de potencias como LPs. Un tipo concreto de LP es la distribución de Pareto. Una variable aleatoria con distribución de Pareto $X \sim \text{Pareto}(\alpha, \sigma)$ tiene función de distribución

$$F(x) = 1 - \left(\frac{x}{\sigma}\right)^{-\alpha}, \text{ con } x \geq \sigma > 0. \quad (2)$$

Al parámetro σ se lo denomina de escala y, nuevamente, α es el parámetro de forma. Merece la pena detenernos un momento a explicar más en detalle ambos parámetros. En muchas ocasiones las variables aleatorias sólo muestran este comportamiento en el extremo de su distribución. Lo hacen a partir de un cierto punto de corte. Dicho punto de corte es σ . Además, se dice que tienen una cola ancha, ya que eventos extremos son mucho más probables que en distribuciones exponenciales, normales, etc. Es el exponente α el que determina cómo de ancha es dicha cola, cuanto mayor, menos probables serán los eventos extremos y viceversa. Volveremos a este tema más adelante.

Para poder entender la naturaleza de aquellos fenómenos regidos por este tipo de distribuciones, hay que pararse a mirar algunas características fundamentales. Supongamos de ahora en adelante que X es una variable aleatoria con distribución de Pareto, $X \sim \text{Pareto}(\alpha, \sigma)$. La esperanza viene dada por

$$E[X] = \begin{cases} \infty, & \alpha \leq 1 \\ \frac{\alpha\sigma}{\alpha-1}, & \alpha > 1 \end{cases}. \quad (3)$$

La expresión (3) nos indica que para aquellas muestras que presenten un parámetro de escala menor o igual a uno, su media será infinita. Esto se puede entender como que en tales situaciones la cola es excesivamente ancha, eventos extremos son muy probables y, por lo tanto, la media deja de tener sentido. La varianza de una tal variable viene dada por

$$\text{Var}(X) = \begin{cases} \infty, & \alpha \in (1, 2] \\ \left(\frac{\sigma}{\alpha-1}\right)^2 \frac{\alpha}{\alpha-2}, & \alpha > 2 \end{cases}. \quad (4)$$

En este caso, la varianza ni siquiera existe para parámetros de escala inferiores a 1, es infinita entre 1 y 2, y cobra sentido de ahí en adelante. Otra vez más, este comportamiento viene a recalcar lo comentado anteriormente.

Por poner un ejemplo, pensemos en la magnitud media de los temblores en la zona de San Francisco, conocida por ser de gran actividad debido a la falla de San Andrés. Los sismógrafos

leen numerosos registros diarios que son imperceptibles para una persona, por lo tanto, el valor medio calculado digamos, durante los últimos cinco años, es una medida cuanto menos engañosa. Es de sobra conocido que en cualquier momento puede ocurrir un terremoto en varios órdenes de magnitud más grande que ese valor que hemos obtenido.

Todo lo explicado hasta aquí está íntimamente relacionado con la teoría de valores extremos. No vamos a indagar más allá en este campo ya que no es el tema fundamental de este trabajo. Si el lector se ha quedado con ganas de más, aconsejo la muy interesante lectura de Alfarano y Lux (2010).

1.2. MÉTODO DE ANÁLISIS DE LEYES DE POTENCIAS

Tal y como habíamos adelantado, un aspecto fundamental a la hora de analizar LPs es el ajuste a partir de datos empíricos de los parámetros que las determinan. Son numerosos los artículos donde se trata de establecer un método más o menos riguroso a la hora de llevar a cabo tal cometido. En algunos se discute incluso la propia existencia de las LPs, ver Farmer y Geanakoplos (2008). En este trabajo nos alejaremos de ese debate y pasaremos directamente al análisis. Emplearemos conceptos trabajados con anterioridad, ver Goldstein et al. (2004), Gabaix (2016), Clauset et al. (2009).

La tarea de ajustar datos empíricos a una ley de potencias no es sencilla, ya que normalmente este comportamiento sólo aparece en la cola alta de la distribución para aquellos eventos raros y extremos.

Para la comprensión del procedimiento seguido en este trabajo se irán introduciendo los conceptos necesarios junto con breves desarrollos. En caso del lector querer indagar más a fondo en los detalles se incluyen artículos donde se satisfará su curiosidad.

Procedimiento

En este apartado vamos a ver: cómo se realiza el tratamiento de una muestra empírica para su posterior representación y ajuste; cómo se obtienen los parámetros del ajuste, el parámetro de forma y el de escala; y cómo se calcula la bondad del ajuste para decidir mediante un p – *valor* si no puede rechazarse la hipótesis nula de que la muestra siga una LP a un determinado nivel significativo o si habría que mirar otros modelos estadísticos.

Empezaremos por el tratamiento de la muestra para su representación y ajuste. El estudio que realizaremos será rango-tamaño. Dada una variable aleatoria X y una muestra empírica con n valores, los ordenamos de menor a mayor $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. La función de distribución empírica viene dada por

$$F_n(x_i) = P(X \leq x_i) = \frac{1}{n+1} \sum_{j=1}^n I_{[x_j \leq x_i]}. \quad (5)$$

En la expresión anterior, el término $I_{[x_j \leq x_i]}$ hace referencia a la función de identidad cuyo valor es 1 si $x_j \leq x_i$ y 0 en caso contrario. La función de supervivencia empírica (6) se define como la complementaria de la función de distribución.

$$S_n(x_i) = P(X \geq x_i) = 1 - \frac{1}{n+1} \sum_{j=1}^n I_{[x_j \leq x_i]}. \quad (6)$$

Si empleamos el concepto de rango definido para la posición i dentro de una muestra de tamaño n como $rango_i = (n+1) - \sum_{j=1}^n I_{[x_j \leq x_i]}$ tenemos que (6) queda expresada:

$$S_n(x_i) = \frac{rango_i}{n+1}. \quad (7)$$

Si empleamos la función de distribución teórica $F(x)$ expresada en (2) para definir la función de supervivencia teórica obtenemos

$$S(x) = 1 - F(x) = \left(\frac{x}{\sigma}\right)^{-\alpha}, \text{ con } x \geq \sigma > 0. \quad (8)$$

Ahora, por simplicidad y para no incluir más variables, suponemos que los n datos de nuestra muestra siguen una ley de potencias (si no, bastará restringir el siguiente desarrollo a aquellos términos de la cola que lo cumplan). Tenemos que

$$\begin{aligned} F_n(x_i) &\approx F(x_i), & \forall x_i > \sigma \\ S_n(x_i) &\approx S(x_i) \end{aligned} \quad (9)$$

...

$$\ln(rango_i) \approx \ln(n+1) + \alpha \cdot \ln(\sigma) - \alpha \cdot \ln(x_i).$$

De esta manera, en la sección 3 se representarán las muestras empíricas para las diferentes variables económicas estudiadas y ajustadas por los parámetros α y σ que iremos obteniendo.

Parámetro de forma

Para poder obtener el parámetro de forma vamos a emplear el método de máxima verosimilitud. Este método nos permitirá estimar los parámetros de nuestra función de densidad mediante la búsqueda del máximo de la función de verosimilitud.

En primer lugar, necesitaremos la función de densidad. Empleando la función de distribución expresada en (2) tenemos que

$$f(x) = \frac{dF(x)}{dx} = \left(\frac{\alpha}{\sigma}\right) \left(\frac{x}{\sigma}\right)^{-(\alpha+1)}, \text{ con } x \geq \sigma > 0. \quad (10)$$

Dada una muestra formada por elementos $x_i > \sigma$, empleando la función de densidad (10) obtenemos la función de verosimilitud (11). Notar que estamos trabajando por encima del factor de escala σ , de momento desconocido. Nos encargaremos de él más tarde.

$$l(x|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \frac{\alpha \sigma^\alpha}{x_i^{\alpha+1}}. \quad (11)$$

Mediante el uso de (11) sacamos la función de logverosimilitud

$$\begin{aligned} \mathcal{L} = \ln(l(x|\theta)) &= \sum_{i=1}^n \ln(f(x_i|\theta)) \\ &= n \cdot \ln(\alpha) + n\alpha \cdot \ln(\sigma) - (\alpha + 1) \cdot \sum_{i=1}^n \ln(x_i). \end{aligned} \quad (12)$$

Es ahora cuando mediante la ecuación de verosimilitud, $\partial \mathcal{L} / \partial \alpha = 0$, y solucionando para α , obtenemos el EMV, estimador de máxima verosimilitud, del factor de forma

$$\hat{\alpha} = \left[\frac{\sum_{i=1}^n \ln\left(\frac{x_i}{\sigma}\right)}{n} \right]^{-1}. \quad (13)$$

El sombrero encima de la variable indica que es un parámetro estimado. Hay que notar que el cálculo depende del parámetro σ , el cual indicaremos cómo se obtiene en el siguiente apartado. Cómo calcularemos el error asociado al parámetro será abordado en el apartado del test de hipótesis, ya que es en ese instante donde cobrará sentido su obtención.

Parámetro de escala

En la gran mayoría de casos, dada una muestra empírica, únicamente la parte alta de la distribución sigue una LP. Es por tanto de vital importancia determinar el valor adecuado de σ . Era práctica habitual ajustar una Ley de Potencias mediante el reconocimiento visual en un gráfico $\log - \log$ del punto de corte, siendo este, el lugar a partir del cual se podía vislumbrar

una línea recta. Posteriormente, autores como Goldstein et al. (2004), pasan a formalizar el ajuste del parámetro de forma de manera rigurosa mediante dos métodos, el antes expuesto y otro empleando el Criterio de Información Bayesiano. Es Clauset (2009) el primero que da un método riguroso para estimar el parámetro de escala σ . Este método ha sido empleado de ahí en adelante en numerosas publicaciones, por ejemplo, Alstott et al. (2014).

Hay que destacar la importancia de escoger correctamente el valor de σ mediante el siguiente razonamiento: en caso de coger un valor de σ excesivamente pequeño, estaríamos ajustando el parámetro de forma α a un conjunto de datos excesivamente grande, donde muchos no formarían parte del modelo que estamos estudiando; por el contrario, si escogemos un valor demasiado elevado, perderemos precisión en la obtención de α ya que estaríamos descartando datos útiles. Este razonamiento queda claro en la figura 1. Se aprecia claramente que el valor de α estimado se aleja rápidamente del valor original, que es 2,5, a medida que cogemos valores de σ más pequeños que el real, que es 300. Por el contrario, el valor permanece estable en caso de quedarnos largos. Este comportamiento es de vital importancia a la hora de tener en cuenta cómo ajustar una LP. Para más detalles, ver Clauset et al. (2009).

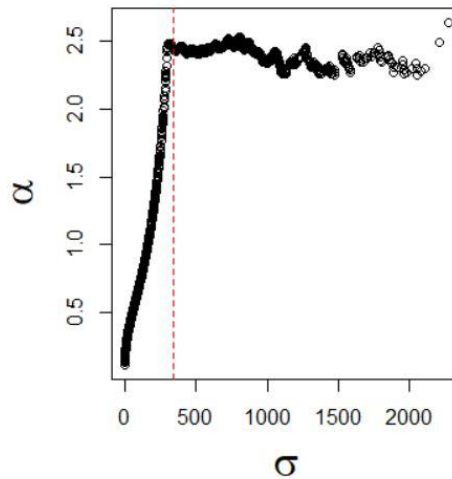


Figura 1. Representación de los valores estimados de α en función del parámetro σ escogido para una muestra sintética con 6000 elementos con la función de distribución indicada en (18).

La manera de obtener el parámetro de escala pasa por escoger el valor σ tal que la función de distribución empírica y la teórica obtenida con los parámetros de ajuste sean lo más similares posibles por encima de σ . Para definir más rigurosamente lo que queremos decir con similares, emplearemos el estadístico de Kolmogorov-Smirnov. El estadístico indicado simplemente mide la máxima distancia entre las funciones de distribución empírica $F_n(x)$ y teórica $F(x)$

$$D = \max_{x \geq \sigma} |F_n(x) - F(x)|. \quad (14)$$

Por lo tanto, $\hat{\sigma}$ será el valor que minimiza D junto con el valor $\hat{\alpha}$ calculado en el apartado anterior. Al igual que con el factor de forma, el cómo se obtiene el error asociado a este parámetro será abordado en el apartado del test de hipótesis.

Test de hipótesis

El procedimiento anterior nos ha permitido ajustar una muestra a una ley de potencias, pero nada dice en relación con cómo de probable es que la función de distribución que originó los datos lo sea. Cuando se llevan a cabo preguntas de este tipo, un método estándar es el test de bondad del ajuste. Lo que se pretende es a partir de la hipótesis nula de nuestra muestra seguir una LP, decidir mediante un $p - \text{valor}$ si hay que rechazarla o no.

A modo de resumen, la idea principal que emplearemos para el test de bondad es la siguiente: generar múltiples muestras sintéticas provenientes de una LP con los parámetros obtenidos por el ajuste inicial de la muestra empírica, ver cuánto se alejan de la función teórica original y comparar con el ajuste de los datos empíricos. Una vez más, para calcular la distancia entre funciones de distribución emplearemos el KS estadístico.

Dada una muestra bajo la hipótesis de seguir una ley de potencias y calculados ya sus mejores parámetros de escala y forma, los pasos a seguir son:

1. Se genera un gran número de muestras sintéticas con dichos parámetros.
2. Se realiza un nuevo ajuste para cada muestra sintética obteniendo para cada una su tupla de parámetros, así como su KS estadístico.
3. El $p - \text{valor}$ es calculado como la fracción de veces donde el KS-estadístico original es menor que el asociado a cada muestra sintética.

La generación de muestras sintéticas sigue el siguiente proceso:

1. Con probabilidad n_{cola}/n , donde n_{cola} determina el número de elementos de la muestra ajustados por una LP, se genera un número aleatorio que sigue la LP con los parámetros estimados.
2. Con probabilidad $1 - n_{cola}/n$ se saca al azar un elemento que no está en la cola y se guarda como valor de la muestra sintética.
3. Se repite el proceso hasta obtener n elementos en total.

A la hora de comprender el resultado arrojado por el $p - \text{valor}$, hay que recordar que este nos indica cuántas veces el ajuste de la muestra empírica a una ley de potencias es mejor que el ajuste de muestras sintéticas que sí se han generado a partir de tal distribución. Por lo tanto, un $p - \text{valor}$ bajo indicaría que nuestra muestra no se asemeja a una LP y por lo tanto habría que rechazar la hipótesis nula. En nuestro caso hemos empleado el mismo umbral que Clauset et al. (2009). Para un $p - \text{valor}$ superior a 0,1 no se rechaza.

El cálculo de la desviación estándar de $\hat{\alpha}$ y $\hat{\sigma}$ se obtendrá mediante las muestras de todos los parámetros obtenidos en todas las muestras sintéticas generadas para el cálculo del $p - \text{valor}$.

Con todos los elementos expuestos en esta sección el lector ya podrá navegar por este trabajo con la confianza de poder comprender todos los resultados arrojados. En caso de querer indagar

más en algunos de los temas tratados, le invito a leer aquellas reseñas indicadas que tan útiles me han sido a la hora de esbozar todo el material.

1.3. LEYES DE POTENCIAS EN LA ECONOMÍA

En numerosas ocasiones, modelos matemáticos creados para entender fenómenos físicos, han sido empleados para el estudio de la economía. Caso singular es el de las LPs, siendo estudiadas primero para dar sentido a fenómenos económicos y extrapolándolas posteriormente a fenómenos físicos, químicos, biológicos, etc.

A lo largo de las últimas décadas algunas teorías han tratado de explicar la existencia de LPs en la economía. Hay que destacar, por un lado, la teoría del crecimiento aleatorio proporcional, Champernowne (1953) y Simon (1955). Ha sido empleada para la simulación del comportamiento de empresas con variaciones estocásticas en su crecimiento. Por el otro, la economía de superestrellas, de manera cualitativa en Rosen (1981), de manera cuantitativa en Gabaix y Landier (2008). En estos artículos se explica la existencia de leyes de potencias en los ingresos dentro de sectores tales como las artes, los deportes y los negocios. Algunos otros fenómenos cuyo comportamiento ha sido modelado por leyes de potencias serían: volatilidad agrupada de los mercados financieros, grandes variaciones de precio en intervalos cortos de los mercados financieros, descuento hiperbólico, la distribución de la riqueza en una sociedad, tamaño de las empresas, etc. Otra relación que también se ha explotado en artículos como Komulainen (2004) y Feldman (2012), es la relación entre LPs y fractales.

En este trabajo, y como ya adelantamos en la introducción, a la hora de estudiar variables económicas, vamos a centrarnos en dos de los aspectos ya mencionados: el comportamiento fractal y la economía de superestrellas.

Fractales

Las leyes de potencias son empleadas en variables que se expresan como libres de escala, ya que individuos de dicha variable pueden tener valores mayores en varios órdenes de magnitud que el resto. Este fenómeno está relacionado con el concepto de autosimilitud, y es aquí donde la teoría fractal tiene que ser tomada en cuenta, ver Takayasu et al. (2000), Takayasu (2009), Gabaix (2008).

La definición de fractal puede remontarse a las publicaciones de Mandelbrot en las décadas de los setenta y ochenta, citando Mandelbrot (1983), se describen los fractales como “una figura geométrica fragmentada o áspera que puede ser dividida en partes, cada cual es una copia de la original”. No indagaremos más profundamente en descripciones técnicas ya que no hay una acepción común. Donde sí vamos a adentrarnos es en el concepto de dimensión fractal.

Cuando se habla de fractales hay un ejemplo muy conocido que identifica muy bien el concepto de dimensión fractal. Hablamos de la paradoja de la línea de costa, observada por primera vez por Richardson en los años cuarenta y analizada posteriormente en Mandelbrot (1983). La paradoja viene a decir que, al medir la longitud de la línea costera, cuanto menor sea la unidad

de medida mayor será la longitud final. Como resultado, si se representa en un gráfico $\log - \log$ la longitud total frente a la unidad de medida, se obtendrá una línea recta, cuya pendiente es la dimensión fractal. Siguiendo Rasband (1990), la dimensión fractal D viene dada por la expresión,

$$n(\epsilon) = k \epsilon^{-D}. \quad (15)$$

El parámetro n es el número mínimo de conjuntos abiertos necesarios para cubrir la figura, ϵ es el tamaño de cada conjunto abierto y k es una constante que depende del tamaño del fractal.

El lector ya habrá notado que la expresión (5) es un tipo de LP, basta comparar con (1). Esta relación puede estudiarse a partir del concepto de autosimilitud.

Siguiendo la línea de razonamiento empleada en Komulainen (2004), un fenómeno autosimilar $A(x)$ es un fenómeno libre de escala que cumple la relación

$$A(\lambda x) = \lambda^s A(x) \quad (16)$$

Donde λ es un factor constante y s es el exponente de escala. Si tomamos ahora una ley de potencias genérica, $y(x) = K x^{-\alpha}$, con K constante, tenemos que

$$y(\lambda x) = K (\lambda x)^{-\alpha} = \lambda^{-\alpha} y(x). \quad (17)$$

De esta manera queda demostrado su comportamiento autosimilar.

Resumiendo todo lo anterior, los fractales son figuras con comportamiento autosimilar caracterizadas por la dimensión fractal, cuyo valor puede ser medido mediante una relación que cumple una ley de potencias.

En este trabajo y en lo referente a este apartado vamos a estudiar las variables económicas que en diversas publicaciones ya se han definido como LPs pero que en este trabajo las estudiaremos a diferentes niveles, buscando comprobar si tienen comportamiento autosimilar y, por lo tanto, podrían entenderse como posibles fractales. Miraremos el número de empleados e ingresos de las principales empresas a nivel global y español, la población en ciudades a nivel global, europeo y español, las variaciones en valor absoluto del Ibex-35 y el cambio de moneda euro- yen y euro-dólar para diferentes escalas temporales.

La economía de las superestrellas

Citando a Rosen (1981), se define a las superestrellas como “un reducido grupo de personas que obtienen enormes cantidades de dinero y dominan las actividades de las que forman parte” (p. 845). Este tipo de entornos se pueden encontrar en distintos sectores. Estamos hablando de, por ejemplo, el mundo de la música, el cinematográfico, el literario, el deportivo, etc. Diferentes

teorías tratan de dar sentido a la aparición de este comportamiento, refiriéndose desde a la relación directa entre ganancias y calidad del producto hasta el talento de los individuos.

En el deporte de alta competición, en numerosas ocasiones nos encontramos o bien a equipos, o bien a deportistas de manera individual, que son acaparadores de los éxitos. En deportes de equipo podríamos poner como ejemplo, dentro del fútbol español, a los equipos Real Madrid CF y FC Barcelona. Ambos han liderado los éxitos deportivos a nivel nacional desde el comienzo de la disciplina en nuestro país. En algunos casos, los salarios están regulados de manera interna por federaciones u organizaciones deportivas. En otros, existe el libre mercado, donde un equipo puede pagar tanto como vea necesario a un jugador por sus servicios. Esta situación se asemeja al estudio realizado en Gabaix (2008). En él se modela la compensación económica de los CEOs de las principales empresas en EE. UU. El modelo resultante sigue una LP. Por ello, nos ha parecido razonable estudiar los salarios en el deporte de la misma manera. Miraremos la NBA, MLB y la EPL a nivel particular y la media salarial por equipos en todas las disciplinas a nivel general.

Pasando a los deportes individuales, si seguimos ahondando en la desproporción por parte de algunos integrantes en la acumulación de éxitos, un ejemplo claro sería el tenis, donde en el lado masculino Federer, Djokovic y Nadal han liderado la disciplina ganando 59 de los últimos 71 (83%) Grand Slam disputados. Este no es un caso único. En el mundo del surf, Kelly Slater y Stephanie Gilmore se han proclamado campeones del mundo un total de 11 y 7 veces respectivamente, siendo claros dominadores de sus disciplinas. Para poder explotar esta particularidad de los deportes individuales, hemos tratado de encontrar una variable económica acorde. Entre algunas de las características que tenía que reunir se encuentran: que sea cuantificable, que los ganadores acumulen mucho más que el resto y, sobre todo, una variable cuyos datos sean fiables y de fácil acceso. Con todo lo anterior, la variable ganancias históricas acumuladas en torneos ha resultado ser una muy buena candidata. Cumple todos los requisitos. En primer lugar, es cuantificable. En segundo lugar, aquellos deportistas dominadores de una disciplina se llevarán mucho más que el resto, ya que en la gran mayoría de casos los premios por resultados crecen muy rápidamente a medida que se llega más lejos en un torneo. Por último, en numerosos deportes existen registros históricos fiables. Estudiaremos esta variable en el tenis, el golf, los e-sports, el bádminton y el snooker. Para todos los casos salvo para el snooker (donde no se han encontrado registros) se hará un estudio tanto para hombres como para mujeres.

Merece la pena destacar que, para esta última variable, premios acumulados, no hemos encontrado ningún estudio previo que la haya relacionado con LPs.

2. METODOLOGÍA DE ESTUDIO

En la sección anterior hemos tratado los temas necesarios para la comprensión del trabajo. En esta sección indagaremos en cómo hemos obtenido los datos, cómo los hemos analizado y con qué variables hemos trabajado.

A la hora de obtener datos no sólo hablaremos de las características que deben cumplir, sino que mencionaremos también ciertos aspectos fundamentales en lo relacionado a la decisión de por qué se han escogido las variables económicas que representan. A la hora de analizar los datos, emplearemos los conceptos presentados en la subsección 1.2 e implementados en el software creado en R. Esto nos permitirá de manera rápida saber qué parámetros mejor ajustarían dicha variable económica a una LP. Además, nos permitirá saber si en efecto una ley de potencias sería un ajuste aceptable o si por el contrario habría que descartarlo y buscar otro modelo estadístico.

2.1. OBTENCIÓN DE LOS DATOS

Tal y como hemos adelantado, en primer lugar, mencionaremos aquellas características que se han tenido en cuenta a la hora de coger un conjunto de datos y darlos por válidos para que de esta manera los resultados sean fiables.

Las propiedades que se han mirado son: la precisión de los datos, la fiabilidad y la actualidad de la fuente y el tamaño de la base de datos. Atendiendo a este último criterio, en el caso de tener un total de valores n pequeño hay que leer los resultados con precaución. Esto se debe a que en tales circunstancias la variable puede superar la prueba del p – *valor* sin ser en verdad una LP. Por lo tanto, cuando n sea pequeño, sólo si el p – *valor* es insuficiente podremos asegurar que no sigue una LP, de lo contrario, simplemente no podremos descartarlo.

Encontrar variables económicas con estas características puede volverse complicado. Mientras que algunas son públicas y por lo tanto de fácil acceso, otras, sobre todo las relacionadas con la economía de superestrellas, no lo son tanto.

Una vez entendidas las características que deben cumplir nuestros datos pasamos a la decisión de qué variables vamos a estudiar.

Tal y como se incluyó en la subsección 1.2, el primer conjunto de variables económicas que vamos a mirar es aquel que autores previos han determinado su comportamiento como LP y que además son susceptibles de ser estudiadas a diferentes niveles. Esto nos permitirá comprobar si dicha variable tiene un comportamiento autosimilar. Tendremos en cuenta dos tipos de variables: las que existen a diferentes niveles geográficos y las que existen en diferentes intervalos temporales. En el primer grupo, miraremos los ingresos por compañías a escala mundial y en España, el número de empleados a escala mundial y en España y, por último, los mayores núcleos urbanos a nivel mundial, europeo y en España. En el segundo grupo, las temporales, miraremos la variación en valor absoluto del Ibex-35 de manera mensual, semanal y diaria y los intercambios Euro-Dollar y Euro-Yen bajo las mismas escalas.

Merece la pena pararnos un segundo y justificar el por qué vamos a trabajar con la variación en valor absoluto de la variable a diferentes escalas temporales. La razón puede encontrarse en Takayasu y Takayasu (2010) donde haciendo referencia al intercambio de monedas, donde dice textualmente *“la simetría arriba-abajo se mantiene de manera universal para escalas temporales cortas de manera general, sin embargo, para unidades mayores de tiempo la distribución de la variación de precio se deforma de manera gradual”* (p. 448). Por lo tanto, en caso de analizar periodos temporales largos estaríamos incurriendo en un sesgo debido a la tendencia a largo plazo del mercado.

En segundo lugar, miraremos variables económicas bajo la teoría de superestrellas. Concretamente nos adentraremos en el mundo del deporte y estudiaremos las variables: salarios y premios acumulados. Las analizaremos para diferentes deportes y en algunas nos adentraremos en un estudio más pormenorizado por posición en el deporte en cuestión y por género.

2.2. ANÁLISIS DE LOS DATOS

En la sección 1 nos habíamos adentrado en los diferentes conceptos teóricos necesarios para el estudio de los datos. Ahora pasaremos a la parte práctica.

En primer lugar, hay que conseguir y manipular los datos. Para ello se han seguido los siguientes pasos:

1. Obtención de la base de datos en el organismo correspondiente.
2. Introducción en una plataforma que nos permita trabajar con ellos.
3. Limpieza, organización y preparación para el programa en R.

En el paso 1 hay que tener en cuenta lo mencionado en la sección 2.1. con la calidad de los datos. En los pasos 2 y 3, se ha trabajado con la herramienta Excel debido a su versatilidad a la hora de poder digerir múltiples formatos de manera sencilla y su rapidez y eficiencia para modificar, resumir y presentar la información al formato requerido. Cualquier otro software sería válido en esta fase.

Una vez tenemos los datos preparados para su uso se introducen en el programa creado en R. Es ahora cuando se hace el estudio de estos y se ve si siguen una LP y con qué parámetros. El algoritmo creado sigue las indicaciones de la sección 1.3. Se resume en:

1. Organización de la muestra y obtención de los mejores parámetros $\{\alpha, \sigma\}$ tal que minimicen el KS estadístico.
2. Creación de conjuntos de datos sintéticos que sigan una ley de potencias con los parámetros obtenidos en 1 y ajustarlos nuevamente con nuevos parámetros.
3. Calcular el p – valor como la fracción de veces que los datos sintéticos muestran un KS estadístico menor que los empíricos.
4. Si el p – valor es menor que 0,1 descartamos la hipótesis nula.

Si el lector quiere ver el código puede dirigirse al apéndice.

Lo primero que habrá que hacer antes de pasar a analizar las variables económicas será comprobar el correcto funcionamiento del programa con el algoritmo presentado. Para ponerlo

a prueba se ha seguido una metodología similar a la empleada por Clauset et al. (2009). Lo que se hace es crear conjuntos de datos sintéticos con distribución

$$f(x) = \begin{cases} C e^{-\alpha(x/\sigma-1)}, & x < \sigma \\ C(x/\sigma)^{-\alpha}, & x \geq \sigma \end{cases} \quad (18)$$

El valor σ actúa como punto de corte bajo el cual la función se comporta como una exponencial y, sobre él, la función se comporta como una LP. Además, esta función en el punto de corte es continua y su derivada es también continua.

El estudio consistirá en un barrido para diferentes valores del número total de datos n , así como diferentes parámetros $\{\alpha, \sigma\}$. Los valores escogidos para n han sido tales que cubran los casos con los que hemos trabajado posteriormente de manera empírica. Hablamos concretamente del intervalo entre 100 y 6000.

2.3. LISTADO DE VARIABLES

Se incluyen a continuación todas las variables estudiadas bajo los criterios ya mencionados. En la subsección 1.2 se explicó cómo todos los datos serían analizados siguiendo una estrategia rango – tamaño. Siguiendo lo expuesto en la subsección 2.1 las dividiremos en dos grupos, aquellas que estudiaremos su comportamiento autosimilar y aquellas que lo haremos dentro del marco de la teoría de superestrellas. En la sección 3 se incluirán los resultados de su análisis.

Volvemos a destacar que en este trabajo pretendemos comprobar si las variables económicas tienen comportamiento autosimilar bajo su estudio a diferentes niveles o bien si dentro de la teoría de superestrellas su cola sigue una LP. Nos encontraremos, por tanto, con variables ya estudiadas como LPs, sobre las cuales haremos un estudio bajo un enfoque diferente o para un mercado diferente, y otras las cuales directamente no han sido escudriñadas bajo la lupa de las LPs.

Variables dentro de la teoría fractal

- i. **Ingresos compañías:** clasificación de las principales empresas por total de ingresos en 2019. Ver para empresas estadounidenses, Axtell (2001), para empresas brasileñas, Da Silva et al. (2018), para empresas chinas Chen et al. (2008). No hemos encontrado estudios previos a nivel España.
 - a. **Mundo:** clasificación a nivel mundial, fuente Fortune (2019). Medido en USD.
 - b. **España:** clasificación a nivel España, fuente Valencia Plaza (2019). Medido en EUR.
- ii. **Número empleados:** clasificación de las principales empresas por total de empleados directos en 2018. En EE. UU. Axtell (2001) y Gabaix (2016). No hemos encontrado estudios previos a nivel España.

- a. **Mundo:** clasificación a nivel mundial, fuente Fortune (2019). Medido en número de personas.
 - b. **España:** clasificación a nivel España, fuente Valencia Plaza (2019). Medido en número de personas.
- iii. **Población ciudades:** clasificación de las ciudades en función del total de su población en 2018. Numerosas publicaciones se han centrado en el comportamiento de la distribución de la población, Sarabia y Prieto (2009), Reed (2001), Gabaix (2016), Gangopadhyay y Basu (2009), Veneri (2016), Decker. et al. (2007).
 - a. **Mundo:** clasificación a nivel mundial, fuente United Nations (2018). Medido en número de personas.
 - b. **Europa:** clasificación a nivel europeo, fuente United Nations (2018). Medido en número de personas.
 - c. **España:** clasificación a nivel España, fuente United Nations (2018). Medido en número de personas.
- iv. **Variación mercado bursátil:** clasificación de las variaciones entre cierres en valor absoluto del Ibex-35. Medido en valor absoluto de la variación porcentual. Ver Bai y Zhu (2010), Levy et al. (1996), Plerou et al. (2004), Kaizoji y Kaizoji (2004).
 - a. **Mensual:** del 04/05/1993 al 02/05/2021, fuente Bolsa de Madrid (2021).
 - b. **Semanal:** del 04/05/1993 al 02/05/2021, fuente Bolsa de Madrid (2021).
 - c. **Diario:** del 04/05/1993 al 06/07/2012, fuente Bolsa de Madrid (2021).
- v. **Variación cambio de moneda:** clasificación de las variaciones entre cierres en valor absoluto. Medido en valor absoluto de la variación porcentual. Ver Yang et al. (2009), Scarlat et al. (2007), Kumar y Maheswaran (2015), Takayasu y Takayasu (2009).
 - a. **Euro-Yen**
 - i. **Mensual:** del 04/05/1993 al 02/05/2021, fuente European Central Bank (2021).
 - ii. **Semanal:** del 04/05/1993 al 02/05/2021, fuente European Central Bank (2021).
 - iii. **Diario:** del 04/05/1993 al 06/07/2021, fuente European Central Bank (2021).
 - b. **Euro-Dólar**
 - i. **Mensual:** del 04/05/1993 al 02/05/2021, fuente European Central Bank (2021).
 - ii. **Semanal:** del 04/05/1993 al 02/05/2021, fuente European Central Bank (2021).
 - iii. **Diario:** del 04/05/1993 al 06/07/2021, fuente European Central Bank (2021).

Variables dentro de la economía de superestrellas

- vi. **Salarios:** clasificación por salarios. Ver Aguinis y O'Boyle (2014), Schneider (2008). En el caso de compensación para los CEOs se han publicado numerosos artículos, Blackwell et al. (2012), Gabaix (2016). En el mundo del deporte ver O'Boyle y Aguinis (2012).
 - a. **Media equipos:** clasificación por salario medio de un equipo dentro de todas las disciplinas en 2019, fuente Sporting Intelligence (2019). Medido en USD. No hemos encontrado estudios previos que tengan en cuenta todas las disciplinas deportivas.

- b. **NBA:** clasificación de los salarios de todos los jugadores de la *National Basketball Association* en 2021, fuente ESPN (2021). Medido en USD. Ver Breslow (2017).
 - c. **MLB:** clasificación de los salarios de todos los jugadores de la *Major League Baseball* en 2021, fuente USA Today (2021). Medido en USD. No hemos encontrado estudios previos relacionados con LPs.
 - d. **EPL:** clasificación de todos los jugadores de la *English Premier League* en 2021, fuente Spotrac (2021). Medido en GBP. No hemos encontrado estudios previos relacionados con LPs.
 - i. **Defensa:** se incluyen defensores y porteros.
 - ii. **Medios:** se incluyen medios.
 - iii. **Delanteros:** se incluyen los jugadores atacantes.
- vii. **Premios acumulados:** clasificación por el histórico de la suma total de premios acumulados dentro de una disciplina concreta. Medido en USD. No hemos encontrado estudios previos relacionados con LPs.
 - a. **Tenis:** clasificación de los jugadores a abril de 2021.
 - i. Hombres, fuente ATP (2021).
 - ii. Mujeres, fuente WTA (2021).
 - b. **E-sports:** clasificación de los jugadores individuales incluyendo todos los e-sports a abril de 2021, fuente E-Sports Earning (2021).
 - i. Hombres.
 - ii. Mujeres.
 - c. **Golf:** clasificación de los jugadores a mayo de 2021, fuente PGA Tour (2021).
 - i. Hombres.
 - ii. Mujeres.
 - d. **Snooker:** clasificación de los jugadores a mayo de 2021, fuente CueTracker (2021).
 - e. **Bádminton:** clasificación de los jugadores a mayo de 2021, fuente BWF (2021).
 - i. Hombres.
 - ii. Mujeres.

3. RESULTADOS

Para poder navegar cómodamente por esta sección el lector debe conocer qué se va a incluir y cómo se va a incluir.

Se ha dividido la sección en dos partes, primero tablas y después figuras. En ambos casos se muestran primero los resultados correspondientes a la comprobación del algoritmo y, después, se incluyen los resultados de las variables económicas, tanto dentro de la teoría fractal como de la economía de superestrellas. Ambas por separado. En estos dos últimos casos, se han incluido valores de la muestra en su conjunto (número de datos, media, desviación estándar y valor máximo) y valores del ajuste realizado. De esta manera el lector puede entender mejor el contexto de la variable.

Otro punto importante es el referente a la indexación de los resultados. En las figuras, a cada gráfica le corresponde un índice. En el caso de la figura 2, que muestra los resultados de la comprobación del algoritmo, los índices se corresponden con cada prueba realizada para una muestra sintética diferente. Se corresponden con los índices de la tabla 1. En el caso de las figuras 3 y 4, los índices son los mismos que empleamos en la subsección 2.3, el listado de variables.

3.1. TABLAS

Comprobación del algoritmo

Tabla 1. Valores de las pruebas realizadas con el programa. Se indica el número de datos de la muestra n , el número de elementos en la cola que siguen una LP, n_{tail} , los parámetros que generaron la muestra α y σ , y, por último, los parámetros calculados por el programa, \widehat{n}_{tail} , $\widehat{\sigma}$, $\widehat{\alpha}$ y el p – valor.

	n	n_{tail}	α	σ	\widehat{n}_{tail}	$\widehat{\sigma}$	$\widehat{\alpha}$	p_{valor}
i.a.	120	100	0,8	1000	99	1200 ± 400	$0,85 \pm 0,11$	0,3
i.b.	240	200	0,8	1000	191	1100 ± 400	$0,81 \pm 0,09$	0,4
i.c.	600	500	0,8	1000	505	1000 ± 300	$0,80 \pm 0,04$	0,5
i.d.	1200	1000	0,8	1000	999	1000 ± 400	$0,80 \pm 0,03$	0,8
i.e.	6000	5000	0,8	1000	5000	1000 ± 300	$0,805 \pm 0,018$	0,6
ii.a.	120	100	2,5	300	75	350 ± 80	$2,6 \pm 0,3$	0,6
ii.b.	240	200	2,5	300	210	310 ± 90	$2,4 \pm 0,2$	0,7
ii.c.	600	500	2,5	300	498	300 ± 60	$2,42 \pm 0,16$	0,7
ii.d.	1200	1000	2,5	300	1000	300 ± 20	$2,48 \pm 0,08$	0,8
ii.e.	6000	5000	2,5	300	4895	310 ± 30	$2,52 \pm 0,04$	0,7

Variables dentro de la teoría fractal

Tabla 2. Parámetros principales en el estudio de cada variable. En orden de aparición tenemos: n número de elementos en la muestra, $\langle x \rangle$ media de los valores en la muestra, $\sigma(\text{std. dev.})$ desviación estándar de la muestra, x_{max} valor máximo de la muestra, $\sigma(x_{min})$ punto de corte donde la muestra se comporta como una ley de potencias, n_{tail} número de observaciones en la cola y porcentaje respecto al total redondeado a las unidades, y p_{valor} el cual nos indica lo probable que la muestra sea una ley de potencias.

FRACTAL – AUTOSIMILAR						
Variable	n	$\langle x \rangle$	$\sigma(\text{std. dev.})$	x_{max}	$\sigma(x_{min})$	$\hat{\alpha}$
GEOGRAFICA						
Ingresos Compañías (EUR)						
Mundial	500	$90 \cdot 10^9$	$67 \cdot 10^9$	$514 \cdot 10^9$	$(8 \pm 2) \cdot 10^9$	$2,1 \pm 0,3$
España	1000	$843 \cdot 10^6$	$172 \cdot 10^6$	$2,2 \cdot 10^9$	$(6 \pm 2) \cdot 10^9$	$1,21 \pm 0,07$
Nº Empleados						
Mundial	500	$205 \cdot 10^3$	$188 \cdot 10^3$	$2,2 \cdot 10^6$	$(2,7 \pm 0,9) \cdot 10^5$	$2,6 \pm 0,7$
España	562	$3,6 \cdot 10^3$	$5,6 \cdot 10^3$	$60,4 \cdot 10^3$	$(4,4 \pm 1,1) \cdot 10^3$	$1,6 \pm 0,2$
Población Ciudad						
Mundial	1000	$1,6 \cdot 10^6$	$2,7 \cdot 10^6$	$31 \cdot 10^6$	$(2,2 \pm 0,5) \cdot 10^6$	$1,43 \pm 0,14$
Europa	500	$4,4 \cdot 10^5$	$7 \cdot 10^5$	$103,8 \cdot 10^5$	$(1,9 \pm 0,5) \cdot 10^5$	$1,46 \pm 0,08$
España	149	$16,9 \cdot 10^4$	$34,4 \cdot 10^4$	$333,5 \cdot 10^4$	$(6,1 \pm 0,6) \cdot 10^4$	$1,29 \pm 0,08$
TEMPORAL						
% Variación Ibex35						
Mensual	340	4,5	4,0	25,2	$8,9 \pm 1,4$	$3,3 \pm 0,6$
Semanal	1475	2,3	2,1	21,2	$4,1 \pm 0,5$	$3,1 \pm 0,3$
Diario	4973	1,1	1,0	14,4	$2,6 \pm 0,4$	$3,2 \pm 0,3$
% Variación EURO-YEN						
Mensual	336	2,5	2,3	16,3	$3,9 \pm 0,8$	$2,6 \pm 0,7$
Semanal	1462	1,2	1,1	12,8	$2,91 \pm 0,15$	$3,8 \pm 0,6$
Diario	4965	0,6	0,6	8,43	$1,67 \pm 0,17$	$3,4 \pm 0,3$
% Variación EURO-DOLLAR						
Mensual	336	2,0	1,8	10,1	$3,7 \pm 0,9$	$3,2 \pm 0,7$
Semanal	1452	1,0	0,8	5,9	$2,8 \pm 0,4$	$4,7 \pm 0,8$
Diario	4950	0,4	0,4	3,8	$1,7 \pm 0,3$	$5,4 \pm 0,9$

Variables en la economía de superestrellas

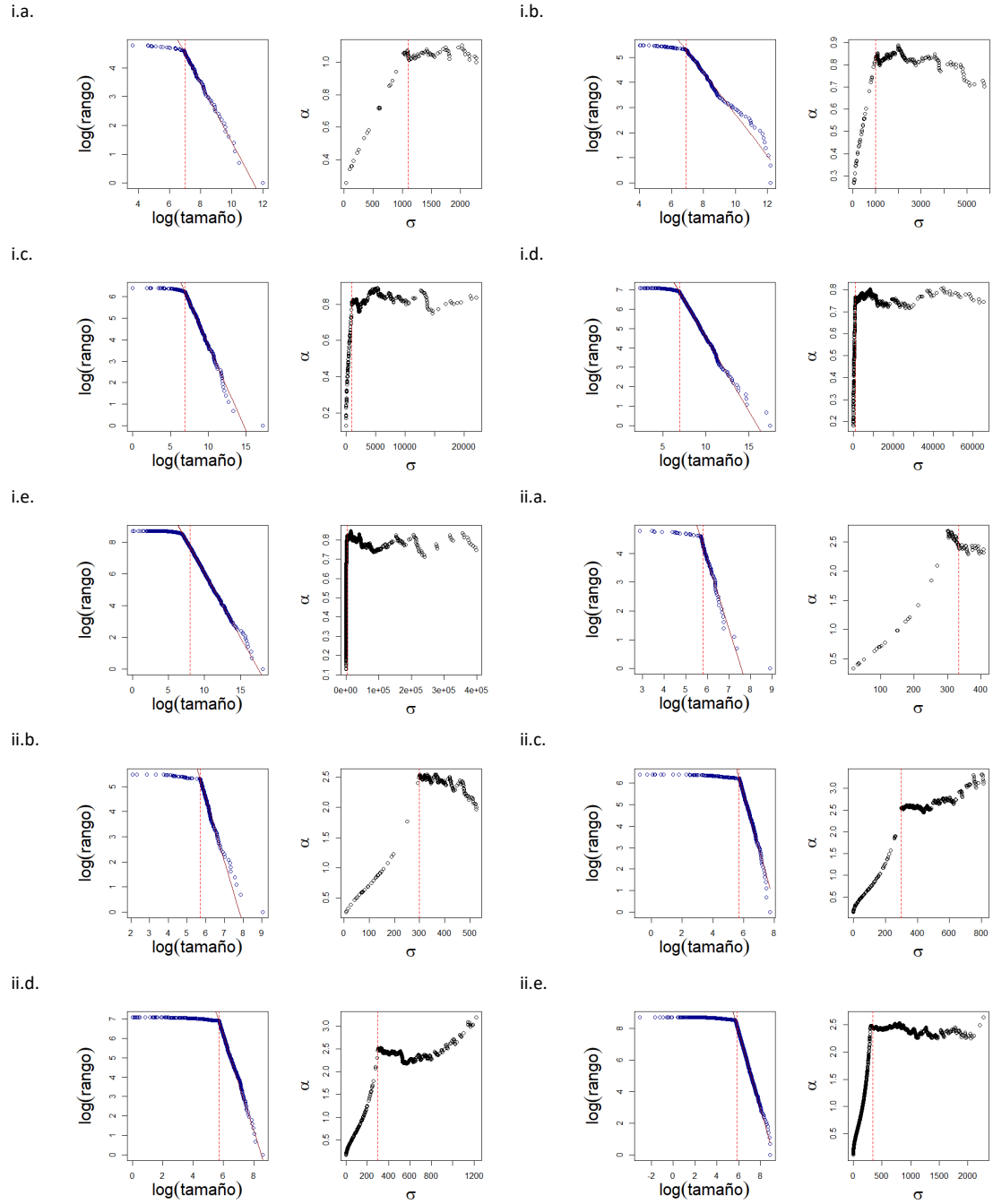
Tabla 3. Parámetros principales en el estudio de cada variable. En orden de aparición tenemos: n número de elementos en la muestra, $\langle x \rangle$ media de los valores en la muestra, $\sigma(\text{std. dev.})$ desviación estándar de la muestra, x_{\max} valor máximo de la muestra, $\sigma(x_{\min})$ punto de corte donde la muestra se comporta como una ley de potencias, n_{tail} número de observaciones en la cola y porcentaje respecto al total redondeado a las unidades, y p_{valor} el cual nos indica lo probable que la muestra sea una ley de potencias.

Variable	n	$\langle x \rangle$	$\sigma(\text{std. dev.})$	x_{\max}	$\sigma(x_{\min})$	$\hat{\alpha}$	n_{tail}	p_{valor}
SALARIOS								
Media equipos	350	$2,6 \cdot 10^6$	$2,5 \cdot 10^6$	$12,3 \cdot 10^6$	$(6,9 \pm 1,1) \cdot 10^6$	$5,2 \pm 0,4$	42(12%)	0
NBA	544	$7,4 \cdot 10^6$	$8,8 \cdot 10^6$	$43 \cdot 10^6$	$(26 \pm 6) \cdot 10^6$	$4,9 \pm 1,5$	41(8%)	0
MLB	898	$4,2 \cdot 10^6$	$6,4 \cdot 10^6$	$37 \cdot 10^6$	$(17,7 \pm 1,9) \cdot 10^6$	$3,4 \pm 0,6$	55(6%)	0
EPL	569	$2,8 \cdot 10^6$	$2,9 \cdot 10^6$	$31,2 \cdot 10^6$	$(5,2 \pm 0,5) \cdot 10^6$	$2,6 \pm 0,3$	76(13%)	0,2
EPL – Defensa	265	$2,4 \cdot 10^6$	$2,1 \cdot 10^6$	$19,5 \cdot 10^6$	$(2,9 \pm 0,3) \cdot 10^6$	$2,4 \pm 0,3$	87(33%)	0,2
EPL – Medios	190	$2,8 \cdot 10^6$	$2,7 \cdot 10^6$	$20,0 \cdot 10^6$	$(4,2 \pm 0,9) \cdot 10^6$	$2,3 \pm 0,4$	78(40%)	0,4
EPL – Delanteros	114	$3,9 \cdot 10^6$	$4,1 \cdot 10^6$	$31,2 \cdot 10^6$	$(3,6 \pm 0,9) \cdot 10^6$	$1,7 \pm 0,3$	43(38%)	0,2
PREMIOS ACUMULADOS								
Tenis H.	500	$6,6 \cdot 10^6$	$11,6 \cdot 10^6$	$147,7 \cdot 10^6$	$(9,7 \pm 2,2) \cdot 10^6$	$1,8 \pm 0,2$	79(16%)	0,9
Tenis M.	500	$4,5 \cdot 10^6$	$7,1 \cdot 10^6$	$94,2 \cdot 10^6$	$(9,1 \pm 1,7) \cdot 10^6$	$1,7 \pm 0,3$	59(12%)	0,4
E-sports H.	1000	$1,1 \cdot 10^5$	$1,0 \cdot 10^5$	$7,0 \cdot 10^6$	$(5,7 \pm 0,6) \cdot 10^5$	$1,56 \pm 0,09$	212(21%)	0,3
E-sports M.	1000	$1,5 \cdot 10^3$	$3,0 \cdot 10^3$	$3,9 \cdot 10^5$	$(7 \pm 5) \cdot 10^3$	$1,2 \pm 0,2$	170(17%)	0,2
Golf H.	642	$9,6 \cdot 10^6$	$25,7 \cdot 10^6$	$120,9 \cdot 10^6$	$(26 \pm 5) \cdot 10^6$	$2,5 \pm 0,5$	58(9%)	0,6
Golf M.	999	$1,2 \cdot 10^6$	$2,5 \cdot 10^6$	$22,6 \cdot 10^6$	$(5 \pm 2) \cdot 10^6$	$2,1 \pm 0,7$	70(7%)	0,3
Snooker	1496	$1,5 \cdot 10^5$	$6,9 \cdot 10^5$	$11,9 \cdot 10^6$	$(6 \pm 2) \cdot 10^5$	$1,1 \pm 0,3$	85(6%)	0,5
Bádminton H.	347	$4,4 \cdot 10^4$	$15,3 \cdot 10^4$	$1,4 \cdot 10^6$	$(8 \pm 3) \cdot 10^5$	$1,08 \pm 0,18$	40(12%)	0,2
Bádminton M.	339	$4,2 \cdot 10^4$	$1,6 \cdot 10^5$	$1,7 \cdot 10^6$	$(6 \pm 2) \cdot 10^3$	$1,05 \pm 0,10$	117(35%)	0,1

3.2. GRÁFICOS

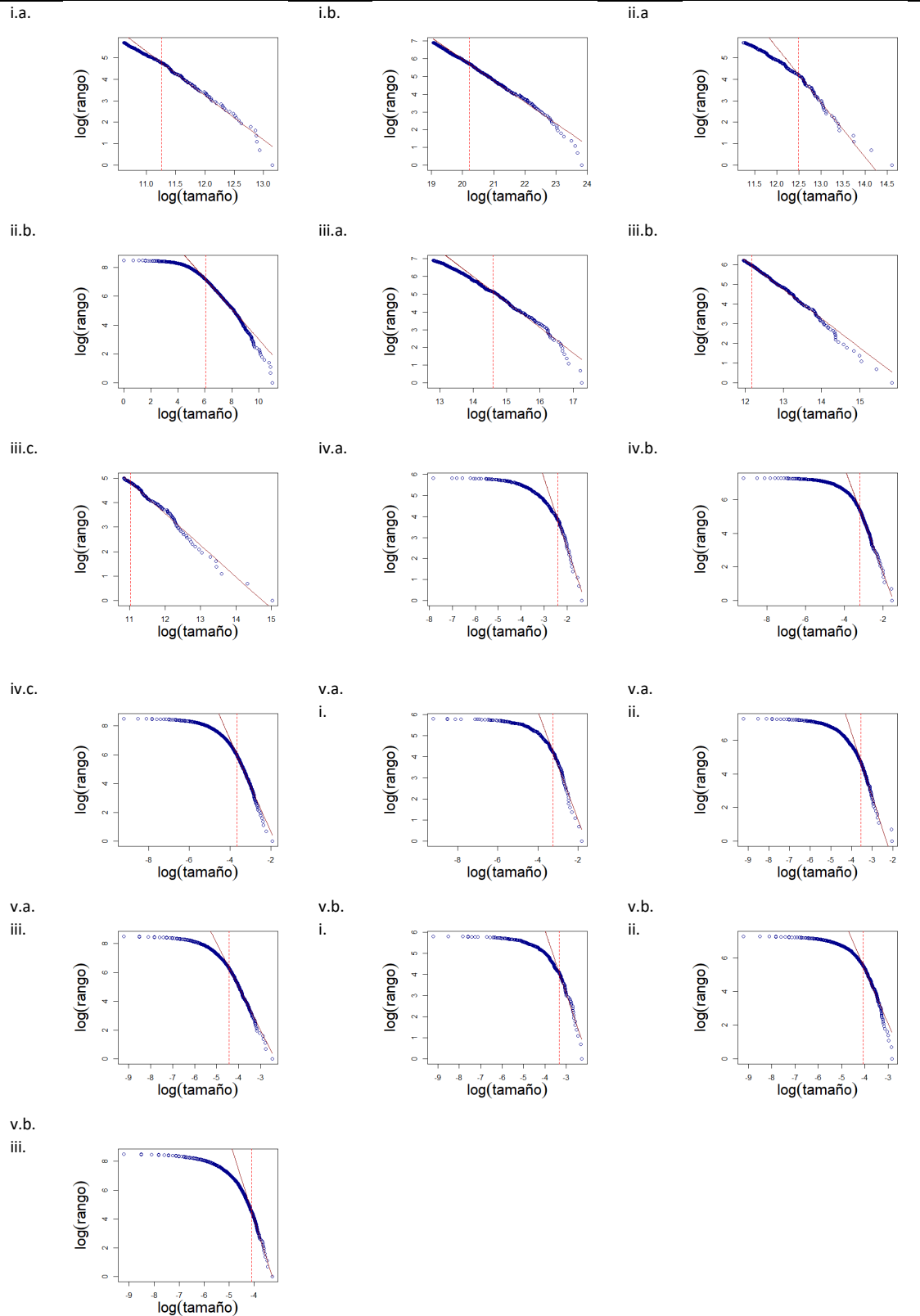
Comprobación del algoritmo

Figura 2. Gráficos para cada una de las pruebas indicadas en la Tabla 1. Para cada caso se han incluido dos gráficas. La de la izquierda es rango-tamaño en escala logarítmica con los datos en círculos azules, el valor de σ que determina la cola en la que se realiza el ajuste como una línea punteada roja y el ajuste realizado como una línea roja sólida. La de la derecha es el comportamiento de $\hat{\alpha}$ frente a $\hat{\sigma}$ para cada tupla posible.



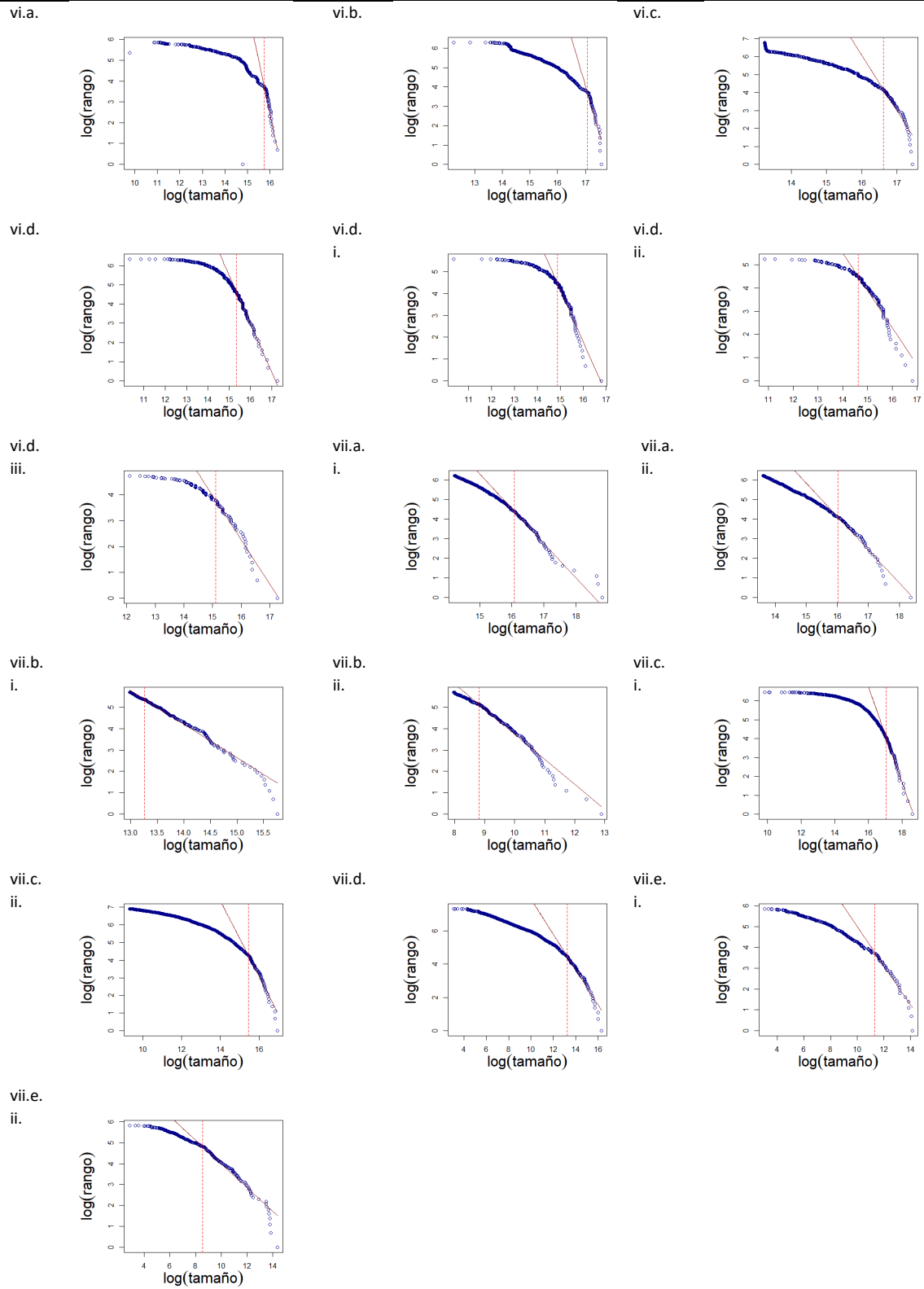
Variables dentro de la teoría fractal

Figura 3. Gráficos rango-tamaño para las variables de la tabla 1 en escala logarítmica con los datos en círculos azules, el valor de σ que determina la cola en la que se realiza el ajuste como una línea punteada roja y el ajuste realizado como una línea roja sólida.



Variables en la economía de superestrellas

Figura 4. Gráficos rango-tamaño para las variables de la tabla 1 en escala logarítmica con los datos en círculos azules, el valor de σ que determina la cola en la que se realiza el ajuste como una línea punteada roja y el ajuste realizado como una línea roja sólida.



4. ANÁLISIS

A la hora de comentar los resultados obtenidos en las diferentes secciones del trabajo seguiremos el orden natural de este. En primer lugar, analizaremos el funcionamiento del programa creado para el análisis de los datos y, en segundo lugar, comentaremos los resultados que arroja el estudio de las diferentes variables económicas.

4.1. ANÁLISIS DEL PROGRAMA

El software creado ha probado tener el comportamiento esperado y ser útil. Para apoyar estas afirmaciones en este apartado estaremos refiriéndonos continuamente a la tabla 1 de la sección 3. Miraremos los siguientes factores: el parámetro de escala, el de forma y las incertidumbres para cada uno en función del tamaño de la muestra.

En todos los casos el programa siempre obtiene un parámetro de forma con una incertidumbre que cubre el valor original y que además muestra comportamiento decreciente. En lo referente al factor de escala, lo primero que hay que mencionar es que obtiene un valor cuyo rango de incertidumbre cubre al valor original. Su incertidumbre hay que recordar que viene dada por el cálculo estadístico a partir de un conjunto de muestras sintéticas generadas a partir del método Bootstrap. Por tanto, el error asociado a la medida siempre se puede modificar aumentando el número de simulaciones generadas con un tamaño apropiado de la muestra. En el caso de tener la potencia computacional suficiente sería recomendable. Los *p – valores* obtenidos para todos los casos son elevados y queda claro que es capaz de indicar con un alto grado de confianza que la muestra analizada sí proviene de una LP.

4.2. ANÁLISIS DE LAS VARIABLES ALEATORIAS

En este apartado haremos una discusión por separado de los resultados obtenidos en las tablas 2 y 3 ya que tenían propósitos distintos.

En la tabla 2 se incluían los resultados para aquellas variables económicas ya estudiadas con anterioridad, pero las cuales hemos mirado desde diferentes escalas, tanto geográfica como temporal.

En el primer grupo de variables, las estudiadas a diferentes escalas geográficas, las dos primeras, ingresos de compañías y número de empleados, muestran un comportamiento similar. Se ve que a diferente escala no tienen el mismo parámetro de forma y por tanto no son autosimilares. Por el contrario, la población agrupada en ciudades sí que muestra un comportamiento autosimilar ya que a diferentes niveles se han obtenido parámetros de forma con una variación entre el menor y el mayor inferior al 12%.

En el segundo grupo de variables, las estudiadas a diferentes escalas geográficas, únicamente la primera, la variación en valor absoluto del IBEX-35, muestra un comportamiento autosimilar. La variación obtenida entre el mayor parámetro de forma y el menor es cercana al 6%. Por el contrario, al comparar los intercambios Euro-Yen y Euro-Dólar vemos que los parámetros de forma muestran un comportamiento errático.

En la tabla 3 se incluían los resultados asociados a las variables económicas dentro del marco de la teoría de superestrellas. En este apartado hemos mirado dos variables económicas dentro del ámbito del deporte, salarios en deportes de equipo y premios acumulados en deportes individuales.

En el estudio de los salarios nos encontramos con un fenómeno curioso ya que hay dos comportamientos bien diferenciados. En primer lugar, la media de salarios por equipo donde se consideran todas las disciplinas deportivas no se comporta como una LP. En la NBA y en la MLB, se aprecia un comportamiento similar. Los ajustes muestran una cola muy alta, con el 6% y el 8% del total de datos respectivamente, arrojando p – valores nulos. Por el contrario, la EPL sí que muestra un comportamiento como LP lo que nos ha permitido realizar un estudio por posiciones. Comparando los factores de forma obtenidos en este último caso se aprecia que, exceptuando a los delanteros, existe un comportamiento autosimilar, ya que entre el resultado obtenido para la liga y las diferentes posiciones hay una diferencia máxima menor al 12%. Cuando miramos a los premios acumulados todas las disciplinas muestran comportamiento como LP. Además, atendiendo al género, hay dos factores a destacar. En primer lugar, todos los parámetros de forma en todas las disciplinas son inferiores en las mujeres que, en los hombres, con un 17% de media. En segundo lugar, exceptuando al bádminton, en todos los casos la cola representa un porcentaje menor del total en las mujeres.

5. CONCLUSIONES

Tras el análisis realizado en la sección 4 hay dos ideas que salen a la luz. **En mercados autoorganizados las LPs aparecen en la cola alta de la distribución y, atendiendo a género en el ámbito del deporte profesional, hay un comportamiento diferente de la élite.** En las siguientes líneas ahondaremos más en estos conceptos siguiendo el mismo orden lógico presentado a lo largo del trabajo.

En lo referente al **comportamiento autosimilar** de las variables económicas estudiadas en la tabla 2, únicamente dos muestran tal propiedad. Estas dos variables, **población y variación del IBEX-35**, a diferencia de los ingresos y el número de empleados, tienen un comportamiento que se puede entender como **libre de regulaciones y restricciones y autoorganizado**. Por otro lado, el hecho de que las variaciones del cambio de monedas Euro-Yen y Euro-Dólar no muestren comportamiento autosimilar va en contraposición a Takayasu y Takayasu (2010). Una mala elección de las escalas temporales podría ser responsable de tal desajuste y, por tanto, indicaría que la variable es susceptible a tal elección. Un estudio más pormenorizado de más cambios de moneda y recorriendo de manera más exhaustiva intervalos de tiempo diferentes sería recomendable.

Pasamos ahora a la tabla 3. En ella se estudian las variables dentro del marco de la **teoría de superestrellas**. Atendiendo a **salarios**, como se comentó en la sección 4, únicamente la EPL no puede ser descartada como una LP. Aquí entran en juego dos factores fundamentales. En la **NBA** existe una franja de salario total por equipo en la que todos los equipos se tienen que mover. Esto hace que existan salarios estándar en función de tu papel en el equipo, antigüedad, etc. En la **MLB** a pesar de no existir un límite salarial sí que existe una penalización proporcional a cuánto se separa el equipo de un cierto máximo establecido. Por tanto, el salario en estos casos **no tiene un comportamiento libre**. Volviendo a la **EPL**, aquí no existe un límite salarial, **no hay regulaciones** y, por lo tanto, los salarios pueden ser firmados a cualquier valor, siendo un **mercado libre y autoorganizado** que puede ser analizado como una **LP en la parte alta de la distribución**. Además, como tampoco existen regulaciones atendiendo a las **posiciones dentro del terreno de juego** la variable muestra un comportamiento **autosimilar**. Este comportamiento es parecido al mostrado por Gabaix (2016) y previamente por Roberts (1956) en relación con la retribución que reciben los CEOs en Estados Unidos. Pasando ahora a la variable **premios acumulados**, aquí no existen restricciones a la hora de decir cuánto un deportista puede llegar a ganar, no hay regulaciones más allá del acceso a los torneos que puedan alterar la **libre competición**. Este razonamiento queda respaldado por los resultados. En todas las modalidades estudiadas se supera el test del p – *valor*, representando la cola donde se realiza el ajuste el 15% de media del total de datos. Esto parece indicar que una **LP** es un **buen modelo** que define la parte alta de la distribución. En la sección 4 se indicó que atendiendo al género de los deportistas y exceptuando al bádminton, tanto el factor de forma como la proporción del número de datos en la cola es mayor en los hombres. De esto se concluye que, en los **hombres**, al ser la cola menor en proporción, **la élite** es un grupo más reducido. Además, el factor de forma más elevado señala que esa élite tiene menos probabilidades de distanciarse con respecto al resto de la competición que en las mujeres. En definitiva, **en la élite son menos y están menos alejados del resto**. Por el lado contrario, lo anterior expuesto también indica que, en el caso de las **mujeres**, **la élite** está formada por **más**, pero **ganan más** con respecto al resto.

5.1. SUGERENCIAS Y MEJORAS

A lo largo de las secciones 4 y 5 se han indicado algunos puntos que debían de ser tratados con especial atención. Merece la pena, por tanto, explicar en un apartado aparte qué consideraciones creemos que deberían tenerse en cuenta a la hora de poder dar mayor valor a las afirmaciones hechas. En primer lugar, en relación con el software desarrollado es recomendable aumentar el número de iteraciones relacionadas con la técnica bootstrap a la hora de calcular el parámetro de escala, así como en el cálculo del $p - valor$. Esto nos permitiría reducir las incertidumbres. En nuestro caso, el trabajo se ha realizado con un ordenador portátil personal con una computación limitada. Esto ha llevado a la elección de un número de iteraciones razonable. También, algunos resultados muestran que ciertas variables podrían ser analizadas desde otros modelos matemáticos. Un posterior estudio podría ser el análisis de otros modelos estadísticos para poder comparar la bondad de los ajustes. Pasando a la sección 5 y, tomando como referencia las ideas presentadas, el paso natural sería: en relación con la idea de mercados autoorganizados y LPs, un estudio con un mayor número de variables económicas; y, en relación con el comportamiento de las élites por género, hacer un estudio más pormenorizado atendiendo a todas las especialidades y ver si puede ser extrapolado a otros ámbitos.

6. REFERENCIAS

- Aguinis, H., & O'Boyle Jr, E. (2014). Star performers in twenty-first century organizations. *Personnel Psychology*, 67(2), 313-350.
- Alfarano, S., & Lux, T. (2010). Extreme value theory as a theoretical background for power law behavior. *Kiel Working Papers* 1648, *Kiel Institute for the World Economy (IfW)*.
- Alstott, J., Bullmore, E., & Plenz, D. (2014). Powerlaw: a Python package for analysis of heavy-tailed distributions. *PloS one*, 9(1), e85777.
- Association of Tennis Professionals (junio de 2021). Singles FedEx ATP Rankings Numerical. Fuente, <https://www.atptour.com/en/media/rankings-and-stats>
- Axtell, R. L. (2001). Zipf distribution of US firm sizes. *Science*, 293(5536), 1818-1820.
- Badminton World Federation (junio de 2021). Prize money by players y Prize money by female players. Fuente, <https://bwfbadminton.com/rankings/>
- Bai, M. Y., & Zhu, H. B. (2010). Power law and multiscaling properties of the Chinese stock market. *Physica A: Statistical Mechanics and its Applications*, 389(9), 1883-1890.
- Blackwell, C., Graefe-Anderson, R., & Vaught, D. (2012). Do Changes in Wealth Inequality Help Explain CEO Compensation Inequality? A Study Using the Pareto Distribution. *A Study Using the Pareto Distribution*.
- Bolsa de Madrid (junio de 2021). Precios históricos del Ibex-35. Fuente, <https://www.bolsamadrid.es/esp/aspx/Indices/InfHistorica.aspx?grupo=IBEX>
- Breslow, J. F. (2017). Salary Inequality in the NBA: Changing Returns to Skill or Wider Skill Distributions? *CMC Senior Theses*. 1645
- Carvalho V. (2008). Aggregate fluctuations and the network structure of intersectoral trade. *Work. Pap., Univ. of Chicago*.
- Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal*, 63(250), 318-351.
- Chen, Q., Zhang, J., & Wang, Y. (2008, October). The Zipf's Law in the Revenue of Top 500 Chinese Companies. *4th International Conference on Wireless Communications, Networking and Mobile Computing* (pp. 1-4). IEEE.
- Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4), 661-703.
- CueTracker (junio de 2021). Most prize money earned by a player up to date. Fuente, <https://cuetracker.net/statistics/prize-money/won/all-time>
- Da Silva, S., Matsushita, R., Giglio, R., & Massena, G. (2018). Granularity of the top 1,000 Brazilian companies. *Physica A: Statistical Mechanics and its Applications*, 512, 68-73.

- Decker, E. H., Kerkhoff, A. J., & Moses, M. E. (2007). Global patterns of city size distributions and their fundamental drivers. *PLoS One*, 2(9), e934.
- Entertainment and Sports Programming Network (junio 2021). NBA players salaries during season 2020-2021. Fuente, http://www.espn.com/nba/salaries/_/page/14/seasontype/5
- E-Sports Earnings (febrero de 2020). Highest overall earnings y highest overall earnings by female players. Fuente, <https://www.esportsearnings.com/players>
- European Central Bank (junio de 2021). Euro foreign Exchange reference rates. Fuente, <https://www.ecb.europa.eu/stats/html/index.en.html>
- Farmer, J. D., & Geanakoplos, J. (2008). Power laws in economics and elsewhere. *Santa Fe: SFI WP; 2008*.
- Feldman, D. P. (2012). Chaos and fractals: an elementary introduction. *Oxford University Press 1st edition*.
- Fortune (junio de 2019). Global 500 by revenues y Global 500 by employees. Fuente, <https://fortune.com/global500/2019>
- Fujiwara, Y. (2004). Zipf law in firm's bankruptcy. *Physica A: Statistical Mechanics and its Applications*, 337(1-2), 219-230.
- Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, 1(1), 255-294.
- Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30(1), 185-206.
- Gangopadhyay, K., & Basu, B. (2009). City size distributions for India and China. *Physica A: Statistical Mechanics and its Applications*, 388(13), 2682-2688.
- Goldstein, M. L., Morris, S. A., & Yen, G. G. (2004). Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 41(2), 255-258.
- Kaizoji, T., & Kaizoji, M. (2004). Power law for ensembles of stock prices. *Physica A: Statistical Mechanics and Its Applications*, 344(1-2), 240-243.
- Komulainen, T. (2004). Self-similarity and power laws. *Complex systems-science on the edge of chaos (Report 145)*.
- Kumar, D., & Maheswaran, S. (2015). Long memory in Indian exchange rates: an application of power-law scaling analysis. *Macroeconomics and Finance in Emerging Market Economies*, 8(1-2), 90-107.
- Levy, M., Solomon, S., & Ram, G. (1996). Dynamical explanation for the emergence of power law in a stock market model. *International Journal of Modern Physics C*, 7(01), 65-72.

- Mandelbrot, B. B., & Mandelbrot, B. B. (1982). The fractal geometry of nature. *Times Books; 2nd edition*.
- O'Boyle Jr, E., & Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology*, 65(1), 79-119.
- Okuyama, K., Takayasu, M., & Takayasu, H. (1999). Zipf's law in income distribution of companies. *Physica A: Statistical Mechanics and its Applications*, 269(1), 125-131.
- Pareto, V. (1896). La curva delle entrate e le osservazioni del prof. Edgeworth. *Giornale degli economisti*, 439-448.
- Plerou, V., Gopikrishnan, P., Gabaix, X., & Stanley, H. E. (2004). On the origin of power-law fluctuations in stock prices. *Quantitative Finance*, 4, C11-C15.
- Professional Golf Association Tour (junio de 2021). Career money leaders up to date. Fuente, <https://www.pgatour.com/stats/stat.110.html>
- Rasband, S. N. (1990). Chaotic dynamics of nonlinear systems. John Wiley & Sons, New York.
- Reed, W. J. (2001). The Pareto, Zipf and other power laws. *Economics letters*, 74(1), 15-19.
- Sarabia, J. M., & Prieto, F. (2009). The Pareto-positive stable distribution: A new descriptive model for city size data. *Physica A: Statistical Mechanics and its Applications*, 388(19), 4179-4191.
- Scarlat, E. I., Stan, C., & Cristescu, C. P. (2007). Self-similar characteristics of the currency exchange rate in an economy in transition. *Physica A: Statistical Mechanics and its Applications*, 379(1), 188-198.
- Schneider, P. A. (2008). A Comparative Entropy Analysis of the Distribution of Wages and Salaries. *The Annual UMass Amherst/NSSR Workshop*.
- Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4), 425-440.
- Sporting Intelligence (junio de 2019). Global sports salaries survey 2019. Fuente, <https://www.globalsportssalaries.com/GSSS%202019.pdf>
- Spotrac (junio de 2021). English Premier League annual salary rankings season 2020-2021. Fuente, <https://www.spotrac.com/epl/rankings/>
- Spotrac (junio de 2021). National Football Association annual salary rankings season 2020-2021. Fuente, <https://www.spotrac.com/nfl/rankings/cash/>
- Takayasu, H., Takayasu, M., Okazaki, M. P., Marumo, K., & Shimizu, T. (2000). Fractal properties in economics. *Papers cond-mat/0008057, arXiv.org*.
- Takayasu, M., & Takayasu, H. (2009). Fractals and economics. *Complex Systems in Finance and Econometrics*, (pp. 444-463).
- United Nations (febrero de 2018). Population of Urban Agglomerations with 300,000 Inhabitants or More in 2018. Fuente, <https://population.un.org/wup/Download/>

- United States of America Today (junio 2021). Major league baseball salaries 2021. Fuente, <https://databases.usatoday.com/mlb-salaries/page/39/>
- Valencia Plaza (febrero de 2019). Ranking de las empresas en España. Fuente, http://epoca1.valenciaplaza.com/nacionalrankings/lista/buscar?limite_pagina=1000
- Veneri, P. (2016). City size distribution across the OECD: Does the definition of cities matter? *Computers, Environment and Urban Systems*, 59, 86-94.
- Women's Tennis Association (junio de 2021). Careers prize money. Fuente, <https://www.wtatennis.com/press-center>
- Yang, Y., Wang, J., Yang, H., & Mang, J. (2009). Visibility graph approach to exchange rate series. *Physica A: Statistical Mechanics and its Applications*, 388(20), 4431-4437.

7. APÉNDICE

Se incluye a continuación el código del software en R empleado para el análisis de las muestras.

```
function_test_power_law <- function(Muestra) {

  # Function that obtains the best fit of a sample to a Power - Law and its p-value
  # Input: The sample contained in a vector
  # Output: Main parameters to be considered and plots of the sample and its fit

  ITERACIONES = 10
  PLOT_TYPE = 1
  MIN_N_TAIL = 40

  Fpareto <- function(mdat, p) {
    1 - (mdat / p[2])^(-p[1]) }

  Fempirica <- function(position, n_length) {
    position / (n_length + 1) }

  Muestra <- sort(Muestra, decreasing = FALSE)
  n <- length (Muestra)
  pos = 0      # It indicates the position where KS is minimized
  pvalorfinal = 0 # Value by default

  X_min_test <- c(1 : (n - MIN_N_TAIL)) * 0
  alpha_test <- c(1 : (n-MIN_N_TAIL)) * 0
  KS_test <- c(1 : (n-MIN_N_TAIL)) * 0

  for(l in 1 : (n-MIN_N_TAIL)) {

    Resample = Muestra[l : n]
    posicion <- rank(Resample, ties.method = "max")
    X_min_test[l] = Resample[1]
    alpha_test[l] = (n - l) / sum(log(Resample / X_min_test[l]))
    p <- c(alpha_test[l], X_min_test[l])

    KS_test[l] = max(abs(Fempirica(posicion, length(Resample)) - Fpareto(Resample, p)))

  }

  pos = which.min(KS_test)
  alpha = alpha_test[pos]
  sigma = X_min_test[pos]
  KS = min(KS_test)

  #----- SEMI-PARAMETRIC BOOTSTRAP -----
```

```

#
# Procedure:
#
# 1- Generate "m" synthetic datasets as follows:
#     Take random x from 0 to 1
#     If  $x > \text{pos}/n$  -> We generate data from power law
#     If  $x < \text{pos}/n$  -> Non-parametric Bootstrap
# 2- Calculate new alpha/sigmas for each data set
# 3- Obtain KS for all data sets
# 4- Obtain p-value and check if  $p > 0.1$  (if benevolent 0.05)
# 5*- If p-value not found, check biyection method with less data
#
# Notes:
# 1- Parametric generation of data:
#     Inverse procedure -> Cuantil (Faustino's procedure)
#     Rejection method -> Similar to MonteCarlo: not done here, too costly
# 2- Code efficiency
#     There is work to be done related to reduction of loops

pvalor <- 0

for (i in 1 : ITERACIONES) {

  Samplesim <- c(1 : n) * 0

  for(l in 1 : n) {
    if(runif(1) > (pos / n)) {
      Samplesim[l] = sigma * (1 - runif(1))(-1 / alpha)
    } else {
      Samplesim[l] = runif(1) * sigma
    }
  }

  Samplesim <- sort(Samplesim, decreasing = FALSE)

  X_min_test2 <- c(1 : (n - 40)) * 0
  alpha_test2 <- c(1 : (n - 40)) * 0
  KS_test2 <- c(1 : (n - 40)) * 0

  for(q in 1 : (n - 40)) {
    Resample = Samplesim[q : n]
    posicion <- rank(Resample, ties.method = "max")
    X_min_test2[q] = Resample[1]
    alpha_test2[q] = (n - q) / sum(log(Resample / X_min_test2[q]))
    p <- c(alpha_test2[q], X_min_test2[q])

    KS_test2[q] = max(abs(Fempirica(posicion, length(Resample)) - Fpareto(Resample, p)))
  }

  if(min(KS_test2) > KS) { pvalor <- pvalor + 1 }

}

```



```

pvalorfinal <- pvalor / ITERACIONES

#----- NON-PARAMETRIC BOOTSTRAP -----
#
# We are willing to obtain the standard deviation of X_min and alpha
# Note: if p-value not sufficient - SKIP
#
# Procedure:
# 1- Resample data with substitution
# 2- Using phase 1, get alpha and X_min
# 3- Repeat i-times and obtain deviation

Set_sigma <- c(1 : ITERACIONES) * 0
Set_alpha <- c(1 : ITERACIONES) * 0

for(i in 1 : ITERACIONES) {

  Resample2 = Muestra[ceiling(runif(n) * n)]

  Resample2 <- sort(Resample2, decreasing = FALSE)
  pos2 <- rank(Resample2, ties.method = "max")

  temp_sigma <- c(1 : (n - 40))
  temp_alpha <- c(1 : (n - 40))
  temp_KS <- c(1 : (n - 40))

  for(l in 1 : (n - 40)) {
    Resample3 = Resample2[l : n]
    posicion <- rank(Resample3, ties.method = "max")
    temp_sigma[l] = Resample3[1]
    temp_alpha[l] = (n - l) / sum(log(Resample3 / temp_sigma[l]))
    p <- c(temp_alpha[l], temp_sigma[l])

    temp_KS[l] = max(abs(Fempirica(posicion, length(Resample3)) - Fpareto(Resample3, p)))
  }

  temp_p = which.min(temp_KS)
  Set_sigma[i] = temp_sigma[temp_p]
  Set_alpha[i] = temp_alpha[temp_p]
}

#----- RESULTS -----
message("----- FINAL RESULTS: -----")

message("n: ", length(Muestra))

message("<x>: ", mean(Muestra))
message("std dev: ", sd(Muestra))
message("x_max: ", Muestra[length(Muestra)])

```

```

message("sigma: ", sigma, "+-", sd(Set_sigma))
message("alpha: ", alpha, "+-", sd(Set_alpha))
message("n_tail: ", n-pos)
message("KS: ", KS)

message("p-valor: ", pvalorfinal)

message("-----")

rango = rev(rank(Muestra, ties.method = "first"))
tamaño = Muestra

if (PLOT_TYPE == 1) {

  par(mar=c(5,6,4,1)+.1)
  plot(log(tamaño), log(rango), pch = 1, col = "darkblue", xlab = expression(log(tamaño)),
ylab = expression(log(rango)), cex.lab = 2.)
  lines(log(tamaño), log(n - pos + 1) + alpha * log(sigma) - alpha * log(tamaño), col =
"darkred")
  abline(v = log(sigma), col = "red", lty = 2)

}else if(PLOT_TYPE == 2) {

  par(mar=c(5,6,4,1)+.1)
  par(mfrow = c(1, 2))
  plot(log(tamaño), log(rango), pch = 1, col = "darkblue", xlab = expression(log(tamaño)),
ylab = expression(log(rango)), cex.lab = 2.)
  lines(log(tamaño), log(n - pos + 1) + alpha * log(sigma) - alpha * log(tamaño), col =
"darkred")
  abline(v = log(sigma), col = "red", lty = 2)
  plot(X_min_test, alpha_test, xlab = expression(paste(sigma)), ylab =
expression(paste(alpha)), cex.lab = 2.)
  abline(v = sigma, col = "red", lty = 2)

}else if(PLOT_TYPE == 3){

  par(mar=c(5,6,4,1)+.1)
  par(mfrow = c(1, 2))
  plot(log(tamaño), log(rango), pch = 1, col = "darkblue", xlab = expression(log(tamaño)),
ylab = expression(log(rango)), cex.lab = 2.)
  lines(log(tamaño), log(n - pos + 1) + alpha * log(sigma) - alpha * log(tamaño), col =
"darkred")
  size = Muestra[pos : length(Muestra)]
  rank = rev(rank(size, ties.method="first"))
  plot(log(size), log(rank), pch = 1, col = "darkblue", xlab = expression(log(tamaño)), ylab =
expression(log(rango)), cex.lab = 2.)
  lines(log(size), log(n - pos + 1) + alpha * log(sigma) - alpha * log(size), col = "darkred")

}
}

```